

PATENT APPLICATION

TECHNIQUES FOR IMPROVING TELEPHONE AUDIO QUALITY

Inventors: Richard Hodges of
Oakland, California
United States citizen

David Roberts of
San Francisco, California
United States citizen

Keith McMillen
Berkeley, California
United States citizen

Leif Claesson
Oakland, California
Citizen of Sweden

Sreevathsa G. Sreenivasa Reddy of
El Cerrito, California
Citizen of India

Keith Edwards of
Antioch, California
United States citizen

Assignee: Octiv, Inc. of Berkeley, California

BEYER WEAVER & THOMAS, LLP
P.O. Box 778
Berkeley, California 94704-0778
(510) 843-6200

TECHNIQUES FOR IMPROVING TELEPHONE AUDIO QUALITY

RELATED APPLICATION DATA

[0001] The present application claims priority from U.S. Provisional Patent Application No. 60/422,430 for EXTERNAL AUDIO PROCESSOR FOR IMPROVING PHONE AUDIO QUALITY filed October 31, 2002, the entire disclosure of which is incorporated herein by reference for all purposes.

BACKGROUND OF THE INVENTION

[0002] The present invention relates to audio signal processing and, specifically, to techniques for improving phone audio quality using adaptive volume control and a variety of signal processing techniques.

[0003] The sophistication of teleconferencing equipment and services has steadily increased as this mode of conducting business has become commonplace in the business world. However, as anyone familiar with these technologies can attest, there are significant shortcomings associated with even the most technologically advanced teleconferencing systems. Typically, these shortcomings relate to level problems, spectral imbalances, and background noise.

[0004] For example, a common problem with which most teleconferencing users are familiar relates to imbalance among the relative volume levels associated with the various parties participating in a conference call. That is, because of the different signal levels associated with different phone equipment and/or the relative positions of various speakers with respect to a particular phone, the voices of some participants are too loud, while others

are often imperceptible. This is particularly the case for analog systems (which still comprise a significant portion of the market), although digital systems may also suffer from similar limitations. In addition, the relatively low fidelity of many telephone infrastructure components, and the resulting noise and distortion further negatively affect the intelligibility of reproduced voice signals.

[0005] It is therefore desirable to provide techniques by which volume imbalances in telephony applications may be mitigated. It is also desirable to provide such techniques which deal with other issues such as, for example, spectral imbalances and background noise.

SUMMARY OF THE INVENTION

[0006] According to various embodiments of the present invention, a variety of techniques including volume normalization and processing of audio signals are employed to improve the experience of users of telephone and teleconferencing services and equipment. According to a specific embodiment, the present invention provides a device for use with a near-end telephone. The near-end telephone is operable to generate an outgoing signal directed to a far-end telephone and to receive an incoming signal generated at least in part by the far-end telephone. A first signal processor is operable to dynamically adjust a first signal level associated with the outgoing signal with reference to the first signal level. A second signal processor is operable to dynamically adjust a second signal level associated with the incoming signal with reference to the second signal level. The first and second signal processors are further operable to control a loop gain to inhibit loop instability. According to some embodiments, the first and second signal processors are operable to control the loop gain by decreasing at least one of a first gain associated with the first signal processor and a

second gain associated with the second signal processor with reference to a combined gain which represents at least a portion of the loop gain.

[0007] According to some embodiments, the combined gain includes a loss component determined with reference to the incoming and outgoing signals. According to a specific embodiment, the loss component comprises an estimate of an echo return loss which is determined with reference to a difference signal representative of a difference between a return energy signal corresponding to the incoming signal and an outgoing energy signal corresponding to the outgoing signal. According to an even more specific embodiment, the estimate deemphasizes speech energy in the incoming signal. In a specific implementation, this is achieved by increasing the estimate according to a time constant when the difference signal exceeds the estimate, and adjusting the estimate to match the difference signal when the difference signal drops below the estimate.

[0008] A further understanding of the nature and advantages of the present invention may be realized by reference to the remaining portions of the specification and the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Fig. 1 is a simplified block diagram of a specific embodiment of the invention.

[0010] Fig. 2 is an alternative representation of portions of the block diagram of Fig. 1.

[0011] Fig. 3 illustrates some exemplary waveforms employed by a loss estimation algorithm for use with a specific embodiment of the invention.

[0012] Fig. 4 is a flowchart illustrating the operation of a specific implementation of a loss estimation algorithm.

[0013] Fig. 5 is a flowchart illustrating operation of a specific implementation of a gain arbitration algorithm.

[0014] Fig. 6 is a simplified block diagram of an analog embodiment of the invention.

[0015] Fig. 7 is a simplified block diagram of a digital embodiment of the invention.

DETAILED DESCRIPTION OF SPECIFIC EMBODIMENTS

[0016] Reference will now be made in detail to specific embodiments of the invention including the best modes contemplated by the inventors for carrying out the invention. Examples of these specific embodiments are illustrated in the accompanying drawings. While the invention is described in conjunction with these specific embodiments, it will be understood that it is not intended to limit the invention to the described embodiments. On the contrary, it is intended to cover alternatives, modifications, and equivalents as may be included within the spirit and scope of the invention as defined by the appended claims. In the following description, numerous specific details are set forth in order to provide a thorough understanding of the present invention. The present invention may be practiced without some or all of these specific details. In addition, well known features may not have been described in detail to avoid unnecessarily obscuring the invention.

[0017] A specific embodiment of the present invention will now be described with reference to Fig. 1. The embodiment shown (i.e., unit 100) is employed with a conventional analog phone circuit in which the outgoing and incoming signals between phones 102 and 104 are transmitted simultaneously on a single pair of wires to and from a phone company central office 105. As will be understood, unit 100 includes circuitry (not shown) which provides all necessary voltages and impedance characteristics between the associated phone and the standard phone jack. It will also be understood, that the present invention may be implemented in digital phone circuits with suitable modifications.

[0018] The incoming and outgoing signals (i.e., the receive and transmit signals) are separated and/or combined using any of a variety of techniques including, for example, conventional hybrid circuits 106 and 108. The non-ideal nature of conventional hybrids

results in a certain amount of echo at each end of a call. This echo is dealt with through the use of near-end and far-end echo cancellers 110 and 112 which may be implemented in hardware or software. The echo cancellers employ any of a wide variety of adaptive signal processing techniques to generate estimates of the near-end and far-end echoes with reference to the incoming and outgoing signals. The estimates are then subtracted (at 114 and 116) from the appropriate signal.

[0019] Each echo canceller works in conjunction with a speech detector (118 and 120) which determines whether incoming or outgoing speech is present, and causes the corresponding echo canceller to update or not accordingly. Speech detectors 118 and 120 may receive as inputs the outgoing and incoming signals as well as the respective error estimates generated by echo cancellers 110 and 112.

[0020] Each of the outgoing and incoming signal paths includes a signal processor 122 and 124 in which various aspects of the present invention may be implemented. According to a particular implementation, signal processors 122 and 124 include one or more automatic gain control (AGC) functions which respond to level imbalances substantially in real time. According to more specific embodiments, signal processor 122 and 124 include multi-band processors capable of addressing the spectral imbalances which detract from the intelligibility of the voice signals. According to various and more specific embodiments, the multi-band processing may be implemented as any of two, three, four, or five band processing.

[0021] Depending upon the widely varying output levels of different telephone equipment, the dynamic range of unit 100 may be more or less efficiently employed. Therefore, according to some embodiments, this issue is addressed by appropriately initializing the gain of signal processors 122 and 124. According to one such embodiment, each of processors 122 and 124 includes a relatively static input AGC and one or more

dynamic AGCs. The input AGC for each processor is initialized at appropriate times, e.g., when unit 100 is powered up for outgoing processor 122, or when each new call begins for processor 124. This sets the operating points of the multi-band AGCs of the processors at their optimal points for the current conditions, i.e., the combination of equipment and associated signal levels for the current call.

[0022] In one embodiment, the gain of the input AGC is set with reference to the received signal strength such that the most or all of the dynamic ranges of the dynamic AGCs are available for dynamic gain control during the call. The dynamic AGC then effects gain control in response to the changing signal levels during the call. According to various embodiments, the dynamic AGCs implemented in signal processors 122 and 124 include wideband AGCs, multi-band AGCs, or combinations thereof. In general, a dynamic AGC employed by the present invention responds dynamically to an input signal, increasing its gain according to an attack rate where the input signal level is below some threshold, and decreasing its gain according to a release rate where the input signal level is above the threshold.

[0023] According to some such embodiments, the multi-band AGCs of signal processors 122 and 124 may be implemented in a wide variety of ways according to a wide variety of techniques. According to some embodiments, this multi-band signal processing is implemented according to the techniques described in commonly assigned, copending U.S. Patent Application No. 10/214,944 for DIGITAL SIGNAL PROCESSING TECHNIQUES FOR IMPROVING AUDIO CLARITY AND INTELLIGIBILITY filed August 6, 2002 (Attorney Docket No. OCTVP001X1), the entire disclosure of which is incorporated herein by reference for all purposes.

[0024] It should be noted that the various signal processing blocks described above and below may be implemented in a variety of ways without departing from the invention. For

example, embodiments of the invention are contemplated in which these blocks are implemented in hardware, software, or a combination of both using one or more microprocessors, application-specific integrated circuits, programmable logic devices, etc. It will also be understood that the operation of the various blocks may be governed by elements within each block itself, or a central control, e.g., a microprocessor, and that the blocks are shown separately for clarity.

[0025] The boosting of the incoming and outgoing signals by signal processors 122 and 124 can have a variety of consequences which must be dealt with to ensure the integrity of the transmitted and received voice signals. For example, some phones with teleconferencing capabilities include a feature to mitigate echo by which local microphones are disabled under certain conditions, e.g., when speech is being received. Again, without the appropriate mechanisms in place, amplification of the incoming and outgoing signals by signal processors 122 and 124 could have the unintended effect of triggering the disabling of a microphone at an inappropriate time, i.e., when only background noise is being transmitted. Therefore, according to some embodiments, noise gating mechanisms, e.g., downward expanders, are included in the loop to reduce the gain under conditions where the signals being transmitted correspond to background noise rather than speech. Any of a wide variety of speech detection algorithms (including any described herein) may be employed to gate such mechanisms.

[0026] Moreover, without appropriate mechanisms in place, it could be possible for the overall system loop gain to exceed 1, and for the system to become unstable (e.g., characterized by high-pitched feedback noise at one or both phones). The circumstances under which this might occur include, for example, when the speech at both ends of a conversation is soft, and/or when the output signals from the phones are relatively low.

[0027] According to a particular set of implementations, this loop gain instability issue is addressed. According to one such implementation, a gain arbitration algorithm is employed in which an estimate of the total gain of the system is maintained, and the gain of various system elements is manipulated to try to prevent the loop gain from exceeding one. The positive contributions to the loop gain estimate include the gains associated with AGCs for each signal direction. These gains are known because they are dynamically controlled to effect the level normalization which is the primary function of the system.

[0028] The loop gain estimate also includes negative contributions representing the losses due to the hybrids and echo cancellation blocks on each end of the loop. According to a specific embodiment, the loss represented by one hybrid and the associated echo cancellation block is estimated by calculating a short term average of the energy going out and a short term average of the energy coming back, and taking the difference between those two values. In calculating these short term averages, it is important not to include energy corresponding to speech or other audio which is intended to be passed through.

[0029] Therefore, according to an even more specific embodiment, a technique for estimating this loss is provided which will be described with reference to Fig. 2, a block diagram representing some of the loop gain and loss components in the system of Fig. 1. In this diagram the gain components of signal processors 122 and 124 are represented by input automatic gain controls (AGCs) 202 and 204, downward expanders (DEs) 206 and 208 (i.e., for noise gating), wideband AGCs 210 and 212, multi-band AGCs 214 and 216, and multipliers 218 and 220. The losses attributable to the echo cancellation blocks and their associated hybrids are represented by blocks 222 and 224 and are equal to the difference between the energy going out of the system, i.e., E_o , and the energy returning, i.e., E_r .

[0030] An exemplary plot of E_o and E_r versus time and on a dB scale is shown in Fig. 3. It should be noted that, for the purpose of clarity, the waveforms in Fig. 3 have been adjusted

in time to account for the round trip delays through the system. It should also be noted that some characteristics of the various waveforms may have been emphasized or exaggerated in some respects to more clearly illustrate the operation of the algorithm.

[0031] As would be expected, much of E_r (waveform 302) appears to be an attenuated version of E_o (waveform 304) because it is the “echo” of the originally outgoing sound. It is this portion of E_r which is comparable with E_o to get an estimate of the loss represented by the hybrid and the echo canceller. However, as mentioned above, a portion of E_r (e.g., portion 306) may represent a voice signal from the other side of the call and should therefore not be included in the estimation of the loss.

[0032] Subtracting E_o from E_r results in a signal $E_r - E_o$ (waveform 308) which is representative of the loss associated with the hybrid and the echo canceller except where sound energy is being received from the other end of the phone call (e.g., portion 310). Another phenomenon occurs when the outgoing energy is extremely low, the necessarily non-zero positive nature of the incoming energy resulting in a hump (e.g., portion 312) in the difference waveform. In both of these situations, the difference waveform ends up not accurately reflecting the loss associated with the hybrid and the echo canceller.

[0033] Therefore, according to various embodiments, a “baseline tracking” algorithm is employed to “smooth out” these transient conditions in the difference waveform and to thereby result in a waveform which yields a loss estimate which can more reliably be used in maintaining the total loop gain to be less than unity.

[0034] According to a particular implementation of the baseline tracking algorithm shown in Fig. 4, a loss estimate waveform 314 is compared with the difference waveform 308 (402). When difference waveform 308 drops below loss estimate waveform 314 (404), loss estimate waveform 314 is immediately adjusted to track the difference waveform 308 (406). On the other hand, when difference waveform 308 is above loss estimate waveform

314 (404), e.g., as might occur when E_r contains voice energy from the far end, loss estimate waveform 314 is only gradually increased (408) until it hits difference waveform 308.

According to various embodiments, the rate of increase is determined such that during normal bursts of energy from the other side only minimal change will occur in waveform 314. The loss estimate represented by waveform 314 is provided to the gain arbitration algorithm which tracks the total loop gain for the purpose of preventing situations where the total loop gain exceeds one.

[0035] According to a more specific embodiment, the above-described algorithm may be used to provide reliable far-end speech detection. That is, where difference waveform 308 is more than a certain amount above loss estimate waveform 314 (410), this portion of the E_r waveform is assumed to be carrying speech energy from the far end (412). This turns out to be a reliable speech detection mechanism which may be used to deal with other issues such as, for example, double-talk situations (where both near and far-end speakers are talking).

[0036] An exemplary implementation of a gain arbitration algorithm for use with the present invention will now be described with reference to Fig. 5. It should be understood that the loss estimates employed by this algorithm are not necessarily derived in the manner described above. It should also be noted that a useful gain arbitration algorithm need not account for every loss or gain element in the loop. For example, the contributions of peak limiters (not shown) may be ignored due to their transitory nature.

[0037] Referring once again to Fig. 2, input AGCs 202 and 204 adapt quasi-statically to the conditions of the connection with reference to the maximum energy level likely to occur during the call. The associated gain levels typically remain static unless energy levels during the call exceed expected levels in which case the gain is reduced. On the other hand, the gains for wideband AGCs 210 and 212 and multi-band AGCs 214 and 216 react dynamically

to the changing energy levels during the call. Multipliers 218 and 220 reduce the gain by some fraction, e.g., half, of the corresponding input AGC gain.

[0038] According to one implementation, the gain arbitration algorithm estimates the loop gain G (502) with reference to the gains of input AGCs 202 and 204 (denoted G_{1A} and G_{1B} , respectively), the gains of wideband AGCs 210 and 212 (denoted G_{2A} and G_{2B} , respectively), the losses of multipliers 218 and 220 (denoted G_{1A}/N and G_{1B}/N), and the loss estimates for the hybrids and echo cancellers in each direction (denoted ERL_A and ERL_B). That is, according to this embodiment, the gain which is monitored by this algorithm is given by:

$$G = G_{1A} + G_{2A} - G_{1A}/N - ERL_A + G_{1B} + G_{2B} - G_{1B}/N - ERL_B \quad (1)$$

It should be noted that the value N may represent either a linear or nonlinear relationship between the gains of input AGCs 202 and 204 and their corresponding multipliers 218 and 220, respectively. It should also be noted that in this embodiment, the effects on the average loop gain of multi-band AGCs and the downward expanders are not taken into account. Other embodiments are contemplated in which either or both of these contributions may be included.

[0039] As mentioned above, if the total loop gain is greater than one (or 0 if measured in dB) (504), undesirable noise artifacts, e.g., feedback, will result. Therefore, according to various embodiments, the value of G is maintained below a certain level by the judicious manipulation of the gain levels associated with various ones of the loop's gain/loss elements (506). Although any of the gain elements in the loop may be manipulated individually or in various combinations, according to some implementations, only the gains associated with wideband AGCs 210 and 212 are manipulated.

[0040] According to one such implementation only the gain of one or the other of these wideband AGCs is typically manipulated. The decision as to which of the two is manipulated is made according to which of the corresponding signal paths is not currently being used to transmit speech energy (as determined, for example, by a speech detection algorithm). This is because the gain in that path is not currently needed, i.e., there is no speech energy, and the gain in the other path is there for a reason, i.e., to enhance the intelligibility of the speech being transmitted.

[0041] According to a more specific embodiment, where speech energy exists in both paths, the gain may be equally reduced in both of the wideband AGCs. It should be noted that this condition is relatively unlikely in that the condition in which the loop gain exceeds one typically occurs where one or both ends of the conversation are quiet, i.e., the gains are high when one or both sides are not speaking.

[0042] According to some implementations of a gain arbitration algorithm designed according to the invention, the necessary gain reductions are not achieved in a single step. Rather, gain is gradually removed from one or more loop elements according to a time constant. According to some embodiments, a suitable time constant might be such that it takes 100 to 500 milliseconds to reduce the gain.

[0043] According to some implementations, the condition in which the total loop gain exceeds one may be made less likely by freezing the gain in one or more of the AGCs when no speech is detected. That is, increases in gain are typically needed when speech is very soft, but are not needed when no one is speaking. Using a speech detection algorithm to distinguish between these two conditions and to gate the operation of the AGCs accordingly can therefore help to reduce the number of situations in which the gain arbitration algorithm might come into play. According to a specific embodiment, the gains of both the wideband and multi-band AGCs are frozen using such a mechanism.

[0044] Other embodiments of the invention deal with the requirement that the amplitudes of the two different DTMF tones must be within 6 dB of each other. According to one such embodiment, the multi-band signal processing is implemented to enforce an interdependence between the bands in which the respective tones are included such that this relationship is never violated. According to another embodiment, the widths of the various bands are set to ensure that both tones fall within the same band and are thus subjected to the same dynamic gain control. According to one such embodiment, two-band signal processing is implemented such that the entire range of DTMF frequencies fall within the same band.

[0045] According to a particular embodiment of the invention of Fig. 1, a bypass circuit is provided with unit 100, i.e., relays 132, 134, and twisted pair 136, which is selected when the phone is on hook and when the phone is first picked up to accommodate ringing signals and services (e.g., Caller ID) provided by the phone company with which unit 100 might interfere. When the phone has been detected as being off hook for a sufficient period of time, unit 100 is switched into the circuit.

[0046] While the invention has been particularly shown and described with reference to specific embodiments thereof, it will be understood by those skilled in the art that changes in the form and details of the disclosed embodiments may be made without departing from the spirit or scope of the invention. For example and as mentioned above, many of the basic techniques described herein may be employed in digital phone circuits with appropriate modifications to ensure compatibility with any of the various digital PBX interface specifications provided by a variety of vendors, e.g., Siemens, Toshiba, Nortel, Alcatel, etc. Figs. 6 and 7 are provided to illustrate the difference between exemplary embodiments for use with analog and digital systems, respectively.

[0047] Fig. 6 is a simplified block diagram of an analog phone system using a specific embodiment 600 of the present invention. By contrast, Fig. 7 is a block diagram of an

exemplary digital phone system using another specific embodiment 700 of the present invention. In digital systems the connection to the phone network is 4-wire. Thus, in the embodiment of Fig. 7 there is no need for a hybrid on the phone network side as there is in the embodiment of Fig. 6. And because there is no hybrid-caused echo to cancel, there is no need for an echo canceller on this side. In such an embodiment, functions such as detection of speech energy coming from the phone network side become simpler because the energy of residual echo can be assumed to be zero.

[0048] It will be understood that, as with the embodiment of Fig. 1, the various blocks and signal paths shown in Figs. 6 and 7 are logically distinct, but may be part of a centrally controlled process. Alternatively, some of the blocks in particular embodiments may be distinct both functionally as well as in terms of hardware.

[0049] In addition, although various advantages, aspects, and objects of the present invention have been discussed herein with reference to various embodiments, it will be understood that the scope of the invention should not be limited by reference to such advantages, aspects, and objects. Rather, the scope of the invention should be determined with reference to the appended claims.